



April 3, 2025

VIA ECF

Hon. Ona T. Wang
Daniel Patrick Moynihan
United States Courthouse
500 Pearl Street
New York, NY 10007-1312

Re: *Authors Guild v. OpenAI Inc.*, Case No. 1:23-cv-08292
Alter v. OpenAI Inc., Case No. 1:23-cv-10211
Opposition to Motion to Compel Cross-Production of Custodial Files [ECF No. 345]

Hon. Ona T. Wang

Page 1

Dear Judge Wang:

Plaintiffs' request for cross-production of the custodial files of Michael Lampe, Brad Lightcap, and Wojciech Zaremba is the latest in a series of attempts to circumvent the custodial discovery process. To be clear, OpenAI has no objection in principle to agreeing to reasonable requests to add additional custodians and/or cross-produce custodial files. Indeed, approximately one month ago, it agreed to add a custodian and cross-produce files from yet another custodian. In total, OpenAI is already producing files from 25 custodians and cross-producing files from another four individuals.

But not all of Plaintiffs' requests have been reasonable, as is the case here. Plaintiffs cannot, for example, identify any non-duplicative information in Messrs. Lampe, Lightcap, or Zaremba's possession. Nor can they explain why—assuming the files are as important as Plaintiffs now claim—they neglected to request them sooner. That failure is particularly glaring, given that every document cited in Plaintiffs' motion was produced by OpenAI 10 months ago.

Plaintiffs' apparent approach is to wait to see what discovery becomes available from other cases and then free-ride off of that discovery, regardless of need, in an attempt to circumvent the limits the Court has imposed in this action. They should not be permitted to do so. OpenAI has already agreed to produce documents from 29 individuals. With discovery now well underway, the Court should reject Plaintiffs' efforts to continually expand the scope of discovery without legitimate need. At the very least, Plaintiffs should be required to justify any additional request for custodians and cross-productions with new and relevant information.

I. Plaintiffs fail to show the requested discovery is non-duplicative.

To justify their demand that OpenAI cross-produce files from additional custodians, Plaintiffs “must demonstrate that [those] custodians would provide *unique* relevant information not already obtained.” *Fort Worth Emp. 's Ret. Fund v. J.P. Morgan Chase & Co.*, 297 F.R.D. 99, 107 (S.D.N.Y. 2013) (emphasis in original); *see also* Fed. R. Civ. P. 26(b)(2)(C) (requiring courts to limit discovery that is “unreasonably cumulative or duplicative”). At the December 3, 2024, status conference, the Court expressed its “strong preference” that Plaintiffs consider whether the discovery they already have covers the information they claim to need “before making a motion for more [] custodians or searches or documents.” Dkt. 296 at 133:9-23. Plaintiffs have not heeded that admonition.

Plaintiffs make no attempt to distinguish their currently designated custodians from Messrs. Lampe, Lightcap, or Zaremba, nor do they identify any *unique* information these three individuals might possess. Most obviously, Plaintiffs' primary rationale for the relevance of these individuals is willfulness. But, as the Court has already observed, any purported evidence of willfulness has almost certainly been captured by the discovery already agreed to and produced by OpenAI. *Id.* at 142:19-21 (“[I]f it's relating to willfulness, you're probably going to get it in what you're getting already.”).

Hon. Ona T. Wang

Page 2

Nor do Plaintiffs' other reasons for seeking Messrs. Lampe, Lightcap, or Zaremba's files hold water. For example, they seek Mr. Lampe's files because of his knowledge of regurgitation. But OpenAI has already identified at least three people who are knowledgeable about that issue, including Mr. Lampe's former team leader and former VP of Safety Systems and the head of OpenAI's Privacy Engineering team. And the Court granted Plaintiffs' prior demand for Ilya Sutskever's files based, in part, on Plaintiffs' assertion that he knew about OpenAI's efforts to address "the problem of ChatGPT's regurgitation of copyrighted material." Dkt. 193 at 2. Plaintiffs seek Mr. Lightcap's files because of his supposed knowledge of efforts to acquire access to copyrighted materials. But OpenAI has already designated at least three custodians with direct knowledge, including its VP of Strategy and Operations, its VP of Product & Partnerships, and its Head of Media Partnerships. Finally, Plaintiffs demand Mr. Zaremba's files because of his purported familiarity with using books for training datasets. But OpenAI has designated at least nine custodians knowledgeable about training, including several who specifically investigated the use of books as training data. *See* Dkt. 198 at 2-3.

Worse, Plaintiffs apparently intend to magnify the duplication problem by continuing to seek cumulative documents from even more custodians going forward. In the course of conferring about this dispute, Plaintiffs flatly refused to consider foregoing discovery from any of the other potential custodians they previously demanded, even though they are seeking *the same materials* from those custodians as from Messrs. Lightcap, Lampe, and Zaremba. For example, Plaintiffs seek cross-production of Messrs. Lightcap, Lampe, and Zaremba because of their knowledge of OpenAI's sourcing and use of books, but they previously demanded Shantanu Jain, Chelsea Voss, Jong Wook Kim, Qiming Yuan, and Andrew Mayne for the *same reason*. *See* Dkt. Nos. 193, 235. Yet Plaintiffs have continued to insist they may seek their files, even if OpenAI agrees to the present demand for cross-production. Thus, Messrs. Lightcap, Lampe, and Zaremba's files are duplicative two times over—not only of the files of OpenAI's designated custodians, but also those of additional individuals Plaintiffs may yet seek to designate. That clear overlap warrants denying Plaintiffs' request. *See Fort Worth Emp. 's Ret. Fund*, 297 F.R.D. at 106–07 (denying request to add proposed custodians on the grounds the plaintiffs failed to demonstrate that the additional requested custodians would provide unique relevant information not already obtained).

II. Contrary to Plaintiffs' assertions, their cross-production demand is burdensome.

Plaintiffs principally justify their cross-production demand by pointing to the lack of burden, asserting that cross-production of the requested files is as simple as "click[ing] 'send.'" Dkt. No. 345. But, to state the obvious, cross-producing custodial files in complex, sophisticated litigation matters is not as straightforward as forwarding an email. The custodial files from Messrs. Lampe, Lightcap, and Zaremba are not sitting in a discrete file or folder. To cross-produce their materials, OpenAI would have to, among other things, (1) run searches to identify their documents on OpenAI's eDiscovery platform, (2) determine which of those documents have already been produced to Plaintiffs (*e.g.*, because they were also in the files of a designated custodian) and remove any such documents from the production set, (3) re-process the documents to address variances between the protective orders and ESI orders in the matter in

Hon. Ona T. Wang

Page 3

which the files were originally produced and this matter, (4) ascertain and fulfill any third-party notice obligations, and (5) package the final production volume with new Bates stamps. Producing additional documents, whether by cross-production or otherwise, also burdens OpenAI by expanding the general bounds of discovery. The more documents produced, the more there will be to address in depositions and the attendant preparation, for example.

More significantly, however, Plaintiffs' serial demands for cross-production are an attempt to end-run the limitations the Court has imposed on custodial discovery in this case. After the Court imposed a limit of 600,000 documents in the News cases, *see* Dkt. No. 293 at 4, the parties agreed to work towards the same limit in this case. But because Plaintiffs refused to reduce their demands, OpenAI in February agreed to review a total of 715,000 documents across Plaintiffs' 25 custodians. This is in addition to the documents OpenAI cross-produced from 4 custodians in other cases. The three additional custodians at issue in Plaintiffs' motion would add even more custodial documents to the tally. The Court has recognized that each expansion of custodial discovery brings with it additional burden on the producing party. *See* Dkt. No. 296 at 57:1-2. That burden should not be discounted.

III. Reasonable limits on requests for additional custodians and cross-production are necessary to minimize burden and ensure discovery proceeds efficiently.

Given the stage of discovery and Plaintiffs penchant for making belated demands for duplicative documents, some restrictions are warranted. Judge Illman recently ordered similar limitations in the parallel action in N.D. Cal. *See, e.g.*, Ex. A at 3 (requiring that "any future requests for additional custodians would be granted only upon a showing of new and relevant information as to why granting the request would be appropriate"); Ex. B at 2-3 (ordering the plaintiffs to "limit any future disputes regarding additional custodians to requests based on evidence produced to Plaintiffs on or after November 25, 2024."). In fact, Plaintiffs themselves previously offered not to seek cross-production of custodial files except (a) on the basis of new information learned or (b) in relation to any deponent on whom Plaintiffs are coordinating depositions, until later reneging on that offer without explanation. Dkt. No. 345-1 at 6-7.

The Court should impose that limitation, or one similar to it, with respect to any future demands for custodian designations or cross-productions. Plaintiffs have had the information they claim was necessary to discern the relevance of Messrs. Lampe, Lightcap, and Zaremba for 10 months, but only recently made any attempt to seek their files. In the intervening time, Plaintiffs have sought the custodial files of 8 other witnesses whom they must have deemed more important than these three individuals. *See* Dkt. Nos. 193, 230, 252, 255. Absent the requested limitation, Plaintiffs will likely return (again and again) with yet another set of demands for custodial documents based on long-possessed information. Plaintiffs should not be permitted to do so. At this juncture, requests for additional custodians should be based on newly acquired information or recent developments in the case. Otherwise, the parties will be forced to inefficiently relitigate Plaintiff's repetitive demands for additional custodial files at every future discovery conference, just as they have at nearly every discovery conference to date.

Hon. Ona T. Wang
Page 4

Respectfully,

KEKER, VAN NEST &
PETERS LLP¹

LATHAM & WATKINS
LLP

MORRISON & FOERSTER
LLP

/s/ Christopher S. Sun

/s/ Herman H. Yue

/s/ John R. Lanham

cc: *All Counsel of Record (via ECF)*

¹ All parties whose electronic signatures are included herein have consented to the filing of this document.